阅读伴随词汇学习的词切分：首、尾词素位置概率的不同作用

梁菲菲 [1,2,3] 冯琳琳 [2] 刘瑛 [2] 李馨 [1,2,3] 白学军 [1,2,3]

（[1] 教育部人文社会科学重点研究基地天津师范大学心理与行为研究院，天津，300387）

（[2] 天津师范大学心理学部，天津，300387）

（[3] 学生心理发展与学习天津市高校社会科学实验室，天津，300387）

摘　要 本研究通过两个平行实验，探讨重复学习新词时首、尾词素位置概率信息作用于词切分的变化模式。采用阅读伴随词汇学习范式，将双字假词作为新词，实验一操纵首词素位置概率高低，保证尾词素相同；实验二操纵尾词素位置概率高低，保证首词素相同。采用眼动仪记录大学生阅读时的眼动轨迹。结果显示：(1)首、尾词素位置概率信息的词切分作用随新词在阅读中学习次数的增加而逐步变小，表现出“熟悉性效应”。(2)首词素位置概率信息的“熟悉性效应”表现在回视路径时间、总注视次数两个相对晚期的眼动指标，而尾词素位置概率信息的“熟悉性效应”则从凝视时间开始，到回视路径时间，再持续到总注视时间。结果表明首、尾词素的位置概率信息均作用于阅读伴随词汇学习的词切分，但首词素的作用时程更长，更稳定，支持了首词素在双字词加工中具有优势的观点。

关键词 词素位置概率，词切分，阅读伴随词汇学习，中文阅读

分类号 B842

1 引言

词是阅读的基本加工单位(Bai et al., 2008; Li et al., 2022; Li & Pollatsek, 2020; Radach & Kennedy, 2004; Rayner, 1998, 2009)。在多数拼音文字书写系统中（如英语、德语等），词间空格是一种天然的词切分线索，帮助读者从视觉上进行词切分，促进词汇识别和引导眼跳定位(Clifton et al., 2016; Perea & Acha, 2009)。由于中文阅读无词间空格之类的视觉词切分线索，词切分过程显得更为复杂(白学军 等, 2019; 梁菲菲 等, 2019; Bai et al., 2013; Blythe et al., 2012; Li et al., 2009; Li & Pollatsek, 2020; Zang et al., 2013)。特别是在阅读中伴随学习

收稿日期：2023-05-26

*基金项目：教育部人文社会科学规划一般项目（21YJA190004）。

通讯作者：梁菲菲，E-mail：feifeiliang_329@126.com

新词时，读者只有将新词切分出来，才可能通过自下而上的词汇水平信息以及自上而下的语境水平信息推断新词语义，逐步构建新词表征并纳入心理词典。中文读者在阅读中依据何种线索将新词切分开来？对该问题的回答有助于理解中文阅读伴随词汇学习的词切分机制，为开发高效词汇学习方式提供理论支持[1]。

词素位置概率信息是中文阅读中一种有效的统计学词切分线索。它是指汉字出现在多字词特定位置的概率（如词首、词中、词尾）(连坤予 等, 2021; Liang et al., 2023)。例如，在“各”字构成的 29 个双字词中（如“各位”、“各自”、“各种”等），“各”字均用在词首。那么，该字出现在词首的概率为 100%，其位置线索完全指向词首；在“员”字构成的 47 个双字词中（如“成员”、“演员”、“员工”等），只有在“员工”一词中“员”字用在词首，其余 46 个词“员”均用在词尾，则该字的位置线索指向词尾。位于词内特定位置的汉字提供了一定的词切分信息，例如，“学校定期举办|各种活动丰富同学们的课余生活”，“各”的出现意味着前一词 n–1“举办”的结束，当前词 n“各种”的开始；同样，看到“员”也意味着词 n 的结束，词 n+1 的开始。

Liang 等人(2015, 2017)通过两项实验，试图回答儿童和成人在阅读中伴随学习词汇时是否会利用词素位置概率信息进行词切分。构造双字假词作为新词，同时操纵首词素和尾词素的位置概率高低，形成三个实验条件：一致条件，首词素常用在词首、尾词素常用在词尾，提供与汉字位置概率一致的切分信息（如“挑尔”）；不一致条件，首词素不常用在词首、尾词素不常用在词尾（如“子左”），提供与汉字位置概率不一致的切分信息；平衡条件，首、尾词素用在词首或词尾的概率相当，均在 50%左右（如“皮合”）。将新词嵌入 6 个强限制性语境中供被试阅读，用以形成新词的词汇表征。结果发现，儿童和成人在不一致条件下对新词的凝视时间和总注视时间显著长于一致条件和平衡条件，表明中文读者在阅读中可以利用词素位置概率信息对新词进行切分。Liang 等人在 Li 等人(2009)关于中文词切分和词识别模型基本假设的基础上，尝试解释了词素位置概率信息在中文阅读中的可能作用方式。在读者

[1] 需要说明的是，由于中文阅读在文本呈现方式上的特殊性（词间无清晰的视觉线索，例如空格），词切分和词识别的关系较为复杂。目前的主流观点（中文阅读眼动控制模型，简称 CRM，见 Li & Pollatsek, 2020；中文 E-Z 阅读者模型, 简称 CEZR 模型，见 Yu et al., 2021）认为，中文阅读的词切分和词识别是统一的过程，完成词切分则意味着完成词识别；反过来，一个词被成功识别则意味着这个词被成功切分。本研究将沿用该主流观点，将词切分和词识别看作同一个过程。为方便理解，文中用词切分来表示。

知觉广度范围内所有汉字被平行激活时，汉字的位置信息也同时被激活。汉字处于词内某一位置的概率越高，汉字处于此位置的激活程度也越高。当激活的信息传递到词单元时，如果已激活汉字的位置信息和词单元中该字当前所处的位置一致时，则容易达到阈限，词就容易被识别。反之，当已激活汉字的位置信息和词单元中该字当前所处的位置不一致时，会造成认知冲突。此时，读者需要花费额外时间解决该冲突。由此推断，词素位置概率信息作用于词加工与识别的“字-组-词”分配环节（character-to-word assignment）。尽管 Liang 等人在研究中证实了词素位置概率信息在阅读伴随词汇学习中的词切分作用。然而，研究者为了实现自变量操纵的最大化，同时操纵了首、尾词素的位置概率信息。因此，无法确定究竟是首词素、尾词素的位置概率信息，还是二者均起到词切分作用。

系列研究表明，首、尾词素的特征在中文词汇识别中的作用及加工方式并不完全相同。相关实验证据如下：（1）首词素的视觉复杂性同时作用于词汇识别和眼跳定位，而尾词素的视觉复杂性仅影响词汇识别，但影响显著小于首词素(Ma & Li, 2015)；（2）首、尾词素的字频均影响词汇识别，但尾词素字频的作用受首词素字频制约(Yan et al., 2006)；（3）首词素特征的激活（如语境多样性、语义透明度）发生时程较早，始于 0~100ms；尾词素特征的激活发生时程较晚，始于 100~200ms (Tsang & Zou, 2022; Wang et al., 2017)。由此推断，首词素在双字词识别和加工中具有一定优势。这可能与中文自身的文字特征以及阅读方向有关：由于中文阅读的视觉词汇加工从左至右进行，读者对首词素的加工先于尾词素，使得首词素在词汇识别中起关键作用。在拼音文字阅读中，首字母组合的主导作用还受限于单词的语音形式，即语音形式是由构成单词所有字母的发音从左到右组合而成(Milledge et al., 2022)。基于上述实验证据，研究者以不同的形式将首、尾词素的不同地位纳入词汇识别模型。例如，自我组织词汇习得与识别模型(Self-organizing Lexical Acquisition and Recognition, 简称 SOLAR, 见 Davis, 2001)主张，字母位置的激活程度从词汇左侧向右侧逐步递减。顺序编码模型(Sequential Encoding Regulated by Inputs to Oscillations within Letter Units，简称 SERIOL, 见 Whitney, 2001)也主张，字母的兴奋性输出在整词上呈梯度变化，变化方向是从词首向词尾逐步减弱。基于首、尾词素在中文双字词识别中的不同作用，有必要进一步明确首、尾词素的位置概率信息在阅读伴随词汇学习中的作用机制。

近期两项研究在中文阅读中考察了首、尾词素位置概率信息的作用，但结果完全对立。Liang 等人(2023)在实验 1 中对首词素位置概率信息的作用进行考察。她们操纵了首词素的位置概率信息高低，同时保证尾词素相同，且用在词首、词尾的概率相当（如“湖水/泉水”）；

实验 2 对尾词素位置概率信息的作用进行考察，操纵尾词素的位置概率信息高低，同时保证首词素相同，且用在词首、词尾的概率相当（如“包括/包含”）。结果发现，尾词素而不是首词素的位置概率信息作用于中文阅读的词切分。曹海波等人(2023)在实验 2a 中考察了首、尾词素位置概率信息在高频词阅读中的作用，在实验 2b 中考察了首、尾词素位置概率信息在低频词阅读中的作用。在每个实验中，他们同时操纵首、尾词素的位置概率信息。结果发现，在高频词加工中，首、尾词素位置概率信息均不起作用；而在低频词加工中，首词素而不是尾词素的位置概率信息作用于中文阅读的词切分。两项研究结果不一致的原因有如下两方面：

第一，核心自变量的操纵方式不同。曹海波等人采用经典的 2×2 实验设计，在一个实验中同时操纵首、尾词素的位置概率信息，形成四个实验条件：首高尾高（如“遗憾”）、首高尾低（如“享受”）、首低尾高（如“责任”）、首低尾低（如“想念”）。Liang 等人则是在操纵首词素位置概率高低的同时，保证尾词素相同，且不提供词切分信息。实验 1 包括两个实验条件：首高尾同（如“湖水”），首低尾同（如“泉水”）；实验 2 包括首同尾高（如“包括”）、首同尾低（如“包含”）。由此推断，首、尾词素位置概率信息的作用方式可能会相互影响。若想回答二者如何共同作用于阅读伴随词汇学习的词切分，前提是需要理解二者各自的词切分作用。

第二，目标词的词频范围不同。曹海波等人研究中的高频词范围在 46~56/百万，低频词范围为 1.57~2.37/百万。Liang 等人研究中目标词的平均词频为 38/百万，介于曹海波等人所使用的高、低频词之间，相当于中频词。由此推断，首、尾词素位置概率信息的作用方式可能受词频所调节。这与 Yu 等人(2021)在中文 E-Z 阅读者模型（Chinese E-Z Reader Model）中所主张的基于字、词熟悉性的词切分计算机制相符。具体来说，在阅读知觉广度范围内的字在一定程度均有激活，读者会依据知觉广度内未被识别的汉字、以及所组成词的熟悉性判断接下来哪几个汉字构成一个词，快速进行词切分。词素位置概率信息作为一种基于汉字位于词内特定位置的构词力的统计学信息，在一定程度上可能会影响字、词的熟悉性，进而影响词切分决策。例如，高频词倾向于整词通达，首、尾词素位置概率信息均不起作用；中、低频词倾向于词素通达，此时，首、尾词素位置概率信息开始起作用，但具体的起作用方式有待进一步研究。新词属于极端低频词，更依赖于自下而上的词素表征，那么，首、尾词素的位置概率信息如何作用于阅读伴随词汇学习的词切分？此外，由于阅读伴随词汇学习的主要特征是“累积性”(Joseph et al., 2014; Joseph & Nation, 2018; Pagán & Nation, 2019)，研究

者常将新词嵌入连续几个不同语境中，帮助读者逐步形成新词的词汇表征。在此过程中，新词由不熟悉逐步向熟悉转变，也就是由低频词逐步向中频词、高频词转变。因此，在阅读伴随词汇学习中考察首、尾词素位置概率信息的作用是否相同，将有助于从词频连续变化的视角对上述问题进行回答。

为此，本研究将通过两项实验分别操纵首、尾词素的位置概率信息，首先回答首、尾词素位置概率信息各自在阅读伴随词汇学习词切分中的贡献。实验一操纵首词素位置概率的高低，同时保证尾词素一致且不提供词切分信息。实验二操纵尾词素位置概率的高低，同时保证首词素一致且不提供词切分信息。基于新词类似于低频词，倾向于词素通达的加工方式(Coltheart et al., 2001)，以及首词素在双字词识别中的加工优势(Ma & Li, 2015; Tsang & Zou, 2022; Wang et al., 2017; Yan et al., 2006)，我们预期：首、尾词素位置概率信息在阅读伴随词汇学习中均发挥作用，且首词素位置概率信息的作用更大。其次，将学习次数作为连续变量纳入模型进行分析，试图回答在新词由不认识到不熟悉再到逐步熟悉的过程中，首、尾词素位置概率信息作用的变化方式。基于词频对词素位置概率的调节作用(曹海波 等, 2023)，我们预期：随着新词学习次数的增加，首、尾词素位置概率信息的词切分作用逐步变小。

2 实验 1 首词素位置概率信息在阅读伴随词汇学习中的作用

2.1 实验方法

2.1.1 被试

64 名天津师范大学在校生参加实验。所有被试母语均为汉语，视力或矫正视力正常，均不知晓实验目的。实验结束后给予被试一定报酬。

样本量的选择参照 Liang 等人(2015, 2017)的研究，效应量水平为 0.48，α 水平为 0.01，G*power 计算结果显示，55 名被试为最小样本量。本实验被试数量 64 名，超过最小样本量。

2.1.2 实验设计

采用单因素两水平（首词素位置概率：高、低）的被试内实验设计。此外，将学习次数作为连续变量纳入模型，用以考察首词素位置概率信息加工的“熟悉性效应”。

2.1.3 实验材料

基于 SUBTLEX-CH 语料库(Cai & Brysbaert, 2010)，选择 111 个汉字作为构成新词的词素。其中，位于双字词词首的概率在 85%以上（如“勾”）、50%左右（如“席”）以及 15%以下（如“望”）的汉字均为 37 个。在高词素位置概率条件下（简称“高概率”），新词由词首概率在 85%以上的汉字和 50%左右的汉字组合而成，如“勾席”；在低词素位置概率条件

下（简称“低概率”），新词由词首概率在 15%以下的汉字和 50%左右的汉字组合而成，如“望席”。最终构造 37 对假词作为新词。为保证假词不真实存在，选取不参与正式实验的 15 名大学生根据拼音写出所对应的词语。最后选取被试全未正确写出的 14 对词作为目标词。

高概率条件下首词素用在双字词词首的平均概率为 93.24%（88.5%~100%）；低概率条件下首词素用在双字词词首的平均概率为 8.84%（0~14%）。两个实验条件下目标词的尾词素相同，且用在词首、词尾的概率均在 50%左右（48%~52%）。此外，对两个实验条件下目标词首词素的笔画数（高概率条件：$M = 6.33$，$SD = 1.91$；低概率条件：$M = 7.07$，$SD = 1.94$）和字频（高概率条件：$M = 444$ 次/百万，$SD = 728$ 次/百万；低概率条件：$M = 261$ 次/百万，$SD = 236$ 次/百万）进行匹配。配对样本 t 检验结果显示，两个实验条件下首词素的笔画数和字频均无显著差异，ts ＜ 1，ps ＞ 0.05。

将假词嵌入 6 个强限制性语境的句子，将其描述成被试所熟悉的某一语义类别的新成员（例如，动物、植物、首饰等），每个语义类别设有两个新成员，分别对应高、低首词素位置概率条件。本实验共包含 14 个语义类别，168 个语境。为控制尾词素相同的一对词在同一语义类别中出现时所产生的实验材料间的干扰，创设了 8 个平衡的组块，保证被试在不同语义类别下阅读尾词素相同的一对词。

每个句子长度均为 16 个汉字，将目标词嵌在句子中间，且目标词首词素和尾词素与其相邻汉字均不构成双字词，即不存在词切分歧义的可能性。分别选取不参与正式实验的 10 名大学生对句子的通顺性和难度进行 5 点等级评定，其中“1”代表句子非常不通顺或句子非常简单，“5”代表句子非常通顺或句子非常难。句子通顺性的平均值为 3.93（$SD = 0.76$），难度的平均值为 1.98（$SD = 0.94$），表明句子通顺且容易理解。

在每个新词嵌入的 6 个语境里，随机呈现 1~2 个阅读理解判断题，考察被试是否真正理解句子含义。同时，为避免阅读理解判断题中出现目标假词，影响词素位置概率的加工进程，所有阅读理解题均在第 3 句之后才会呈现。此外，为考察被试对新词语义类别的掌握程度，在读完第 6 个语境后呈现一个语义类别选择题，共包含四个选项：两个选项中的语义类别来自于正式实验，另外两个是填充项。实验材料及实验范式见表 1。

表 1 实验材料及实验范式

学习顺序	句子框架
1	从选材到制作勾席/望席需要经过多重工序。
2	多次地切割使得勾席/望席更加的光彩夺目。
3	很多年轻人认为勾席/望席代表浪漫的爱情。
4	很多人在购买勾席/望席之前要量手指尺寸。
5	新郎和新娘交换勾席/望席戴在对方手指上。
6	为女朋友挑选勾席/望席需要花费很多心思。
语义类别选择题	请问：勾席/望席属于以下哪个类别？
	（a）首饰 （b）动物 （c）杯具 （d）防疫用品

注：勾席为高首词素位置概率条件；望席为低首词素位置概率条件

2.1.4 实验仪器

采用 EyeLink1000 眼动仪，采样率为 1000Hz，屏幕分辨率为 1024×768 像素，刷新率为 120Hz。被试与屏幕之间的距离为 70cm。字体为宋体 18 号，每个汉字大小为 25×25 像素，约呈 0.80° 视角。

2.1.5 实验程序

被试单独施测。首先，进行水平三点校准，平均误差小于 0.25° 。校准成功后，呈现指导语，在被试理解实验要求后进入练习试次。随后进入正式实验。一屏呈现一个句子，阅读完毕后，按“空格键”翻页。回答阅读理解题目时，使用鼠标“左键”选择屏幕中相应的正确答案。在 6 个句子全部阅读完毕后，被试需根据 6 个语境的描述判断所学新词的语义类别，同样使用鼠标“左键”选择正确答案。整个实验持续 30 分钟左右，为缓解被试疲劳，10 分钟左右让被试休息 1~2 分钟。

2.1.6 数据分析

参照前人研究(Liang et al., 2015, 2017, 2023)，选取首次注视时间、凝视时间等反映词汇识别早期的眼动指标以及回视路径时间、总注视时间、回视出比率、总注视次数等反映词汇识别晚期的眼动指标作为因变量。基于 R(R Development Core Team, 2016)语言环境下的线性混合模型（liner mixed model, LMM）、广义线性混合模型（generalized mixed-effects models，GLMMs）和 lme4 数据包(Bates et al., 2023)进行数据分析。对时间类眼动指标进行 log 转换。将首词素位置概率作为固定因素，学习次数作为连续变量，被试、项目作为随机效应纳入模型。采用最大随机效应结构模型，若无法拟合，则采用逐渐递减原则，直至模型拟合成功。

2.2 实验结果

根据以下标准删除数据(白学军 等, 2019; 梁菲菲 等, 2019; Liang et al., 2015, 2017)：（1）注视点持续时间小于 80ms 或大于 1200ms；（2）眼动追踪信号丢失；（3）单个句子注视点少

于 3 个；（4）3 个标准差之外。删除数据占总数据的 0.3%。

阅读理解选择题的平均正确率为 97.77%；语义类别选择题的平均正确率为 95.15%。表明被试在实验过程中认真阅读了实验语句，并习得了新词的语义类别。

高、低概率实验条件下对目标新词的注视情况见图 1，模型结果分析汇总见表 2。

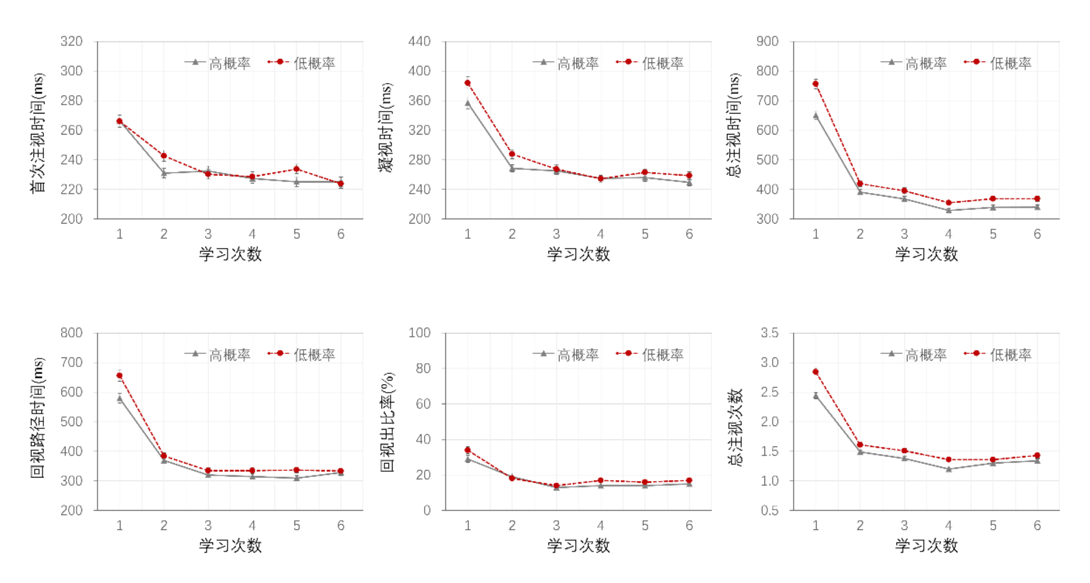


图 1 高、低首词素位置概率条件下新词的注视情况

表 2 不同词素位置概率实验条件下的模型分析汇总结果

	变量	*b*	*SE*	*t/z*	95%CI
首次注视时间	截距	5.49	0.01	457.88	[5.46, 5.51]
	首词素位置概率	–0.01	0.02	–0.62	[–0.04, 0.02]
	学习次数	–0.03	0.005	–5.21	[–0.04, –0.02]
	首词素位置概率×学习次数	0.0002	0.004	0.05	[–0.01, 0.01]
凝视时间	截距	5.67	0.03	177.65	[5.61, 5.73]
	首词素位置概率	–0.04	0.02	–2.11	[–0.08, –0.003]
	学习次数	–0.05	0.01	–8.39	[–0.06, –0.04]
	首词素位置概率×学习次数	0.01	0.01	1.01	[–0.01, 0.02]
总注视时间	截距	6.20	0.05	137.37	[6.12, 6.29]
	首词素位置概率	–0.10	0.02	–4.32	[–0.15, –0.06]
	学习次数	–0.11	0.01	–12.87	[–0.12, –0.09]
	首词素位置概率×学习次数	0.01	0.01	1.37	[–0.004, 0.02]
回视路径时间	截距	6.02	0.04	136.97	[5.94, 6.11]
	首词素位置概率	–0.10	0.03	–3.53	[–0.15, –0.04]
	学习次数	–0.09	0.01	–10.35	[–0.11, –0.07]
	首词素位置概率×学习次数	0.01	0.01	2.04	[0.001, 0.03]
回视出比率	截距	–1.13	0.14	–7.79	[–1.41, –0.84]
	首词素位置概率	–0.21	0.13	–1.61	[–0.46, 0.05]
	学习次数	–0.17	0.03	–5.99	[–0.22, –0.11]

	首词素位置概率×学习次数	0.02	0.03	0.69	[–0.04, 0.09]
总注视次数	截距	2.32	0.09	26.55	[2.15, 2.49]
	首词素位置概率	–0.33	0.05	–6.63	[–0.42, –0.23]
	学习次数	–0.20	0.02	–12.31	[–0.24, –0.17]
	首词素位置概率×学习次数	0.05	0.01	3.97	[0.02, 0.07]

在所有眼动指标分析中，学习次数的主效应均显著（$|t/z|$s $>$ 5.21, ps $<$ 0.001），随着新词学习次数的增加，被试对新词的注视时间逐步缩短，回视比率逐步降低，重复了阅读伴随词汇学习的“累积性”特点。

在首次注视时间分析中，首词素位置概率的主效应，以及与学习次数的交互作用均不显著（$|t|$s $<$ 0.62, ps $>$ 0.05），表明在新词加工的早期阶段，被试对首词素的位置概率信息不敏感。

在凝视时间分析中，首词素位置概率效应显著（$|t| = 2.11, p = 0.03$），高概率条件下目标词的凝视时间显著短于低概率条件，表明在新词加工的相对早期阶段，读者开始加工首词素的位置概率信息；首词素位置概率与学习次数的交互作用不显著（$t = 1.01, p > 0.05$），表明首词素位置概率信息稳定地作用于新词学习的全程。

在回视出比率分析中，首词素位置概率的主效应，以及与学习次数的交互作用均不显著（$|z|$s $<$ 1.61, ps $>$ 0.05）。但在回视路径时间分析中，首词素位置概率的主效应，以及与学习次数的交互作用均显著（$|t|$s $>$ 2.04, ps $<$ 0.05，交互作用见图 2a），被试在高、低概率条件下对目标词前语境回视时间的差异随新词在阅读中学习次数的增加逐步减小。该结果表明，新词首词素位置概率的高低不会影响读者对目标词前语境的回视比率，但是会影响目标词前语境的注视时间。

在总注视时间的分析中，首词素位置概率以及学习次数的主效应均显著（$|t|$s $>$ 4.32, ps$<$ 0.001），但二者的交互作用不显著（$t = 1.37, p > 0.05$）。在总注视次数分析中，首词素位置概率的主效应，以及与学习次数的交互作用均显著（$|t|$s $>$ 3.97, ps $<$ 0.001，交互作用见图 2b），随着新词学习次数的增加，被试在高、低概率条件下对目标词总注视次数的差异随新词在阅读中学习次数的增加逐步减小。

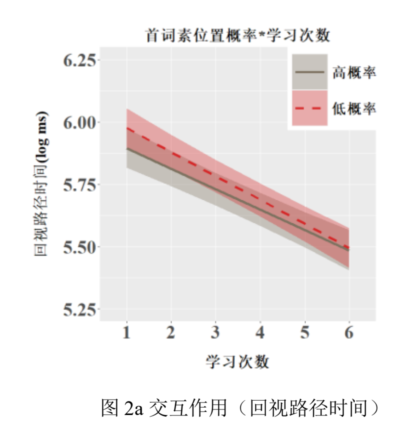


图 2a 交互作用（回视路径时间）

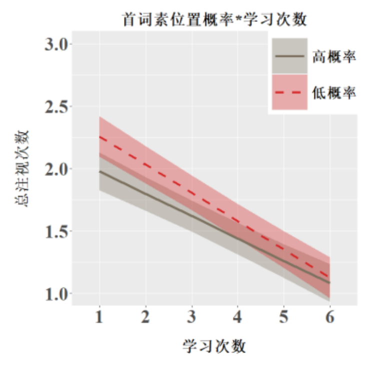


图 2b 交互作用（总注视次数）

2.3 讨论

实验 1 通过操纵新词首词素位置概率的高低，考察了中文读者在阅读伴随词汇学习中是否利用该信息进行词切分。本实验的第一个重要发现是，在反映词汇加工相对晚期的眼动指标中（如凝视时间、回视路径时间、总注视时间以及总注视次数），发现了显著的首词素位置概率效应，即新词首词素常用在词首时，其加工时间显著短于新词首词素不常用在词首时。该结果与本研究的第一个预期相符合，表明首词素位置概率信息作用于阅读伴随词汇学习的词切分。该发现与 Liang 等人(2023)的研究结论不一致，但与曹海波等人(2023)在低频词中关于首词素位置概率的作用结论一致。基于上述两项研究以及本研究中对目标词词频的操纵差异，可以推断，高频词(曹海波 等, 2023)和中频词(Liang et al., 2023)加工中没有出现首词素位置概率效应，而在低频词和新词加工中却存在首词素位置概率效应。在总讨论中，我们将同时结合首、尾词素位置概率信息的结果发现，详细阐释词频调节词素位置概率信息加工的内在机制。

本实验的第二个发现则是，在回视路径时间和总注视次数两个晚期眼动指标上，首词素位置概率和学习次数的交互作用显著。随着新词在阅读中学习次数的增加，首词素位置概率效应逐步减小，最后消失。该结果与本研究的第二个假设一致，表明首词素位置概率信息的词切分作用表现出“熟悉性效应”或“学习效应”。结合阅读伴随词汇学习的累积性特点，首词素位置概率信息加工的“熟悉性效应”实则与新词词汇表征逐步构建，由极端低频词逐步向高频词转变的过程相关。进一步解释为，当新词在阅读中首次出现之前，读者头脑中没有关于新词表征的任何信息，是一个完全意义上的极端低频词；当新词首次在阅读中出现时，读者不得不依据语境以及新词首词素的位置概率信息进行词切分和词识别，在此过程中初步构建新词的形、音、义表征；随着新词在阅读中出现次数的增多，新词表征构建的越来越完善，读者可以在一定程度上依据已存储的新词表征自上而下地词切分，此时将不再依赖于自下而上的首词素位置概率信息，由此表现出新词首词素位置概率效应的逐步消失。

3 实验 2 尾词素位置概率信息在阅读伴随词汇学习中的作用

3.1 实验方法

3.1.1 被试

另选 64 名天津师范大学在校生作为被试。被试选择标准同实验 1。

3.1.2 实验设计

采用单因素两水平（尾词素位置概率：高、低）被试内实验设计。此外，将学习次数作为连续变量纳入模型进行分析，用以考察尾词素位置概率信息加工的“熟悉性效应”。

3.1.3 实验材料

基于 SUBTLEX-CH 语料库(Cai & Brysbaert, 2010)，选择 132 个汉字作为构成新词的词素。其中，位于双字词词尾的概率在 85%以上（如“坛”）、50%左右（如“朴”）以及 15%以下（如“吊”）的汉字均为 44 个。目标词的构造方法同实验 1，在高尾词素位置概率条件下（简称“高概率”），新词的尾词素由词尾概率在 85%以上的汉字构成；在低尾词素位置概率条件下（简称“低概率”），新词的尾词素由词尾概率在 15%以下的汉字构成。两个实验条件下，新词的首词素为同一汉字，用在词首、词尾的概率均在 50%左右。为保证本实验所使用的假词均为“假词”，选取不参与正式实验的 15 名大学生根据拼音写出所对应的词语。最后选取被试全未正确写出首尾词素的 15 对词作为目标词。

两个实验条件下目标词的操纵和匹配方式同实验 1，描述统计见表 3 所示。配对样本 *t* 检验结果显示，两个实验条件下尾词素的笔画数和字频均无显著差异。

表 3 实验材料匹配结果

尾词素位置类型	目标词举例	尾词素平均词素位置概率	首词素平均词素位置概率	尾词素平均笔画数	尾词素平均字频
高概率条件	朴坛	93.5%（86%~100%）	50%（48%~52%）	7.13（2.58）	183（394）
低概率条件	朴吊	8.4%（0~13%）	50%（48%~52%）	7.07（2.32）	254（486）

注：字频的单位是次/百万。

为排除实验句子框架不同对两个实验结果带来的影响，实验 2 采用与实验 1 相同的句子框架。实验材料及实验范式见表 4 所示。

表 4 实验材料及实验范式

学习顺序	句子框架
1	从选材到制作朴坛/朴吊需要经过多重工序。
2	多次地切割使得朴坛/朴吊更加的光彩夺目。
3	很多年轻人认为朴坛/朴吊代表浪漫的爱情。
4	很多人在购买朴坛/朴吊之前要量手指尺寸。
5	新郎和新娘交换朴坛/朴吊戴在对方手指上。
6	为女朋友挑选朴坛/朴吊需要花费很多心思。
语义类别选择题	请问：朴坛/朴吊属于以下哪个类别？
	（a）首饰　（b）动物　（c）杯具　（d）防疫用品

注：朴坛为高尾词素位置概率条件；朴吊为低尾词素位置概率条件

3.1.4 实验仪器和实验程序

同实验 1。

3.2 结果

眼动数据删除标准同实验 1，删除数据占总数据的 0.2%。眼动指标选择和数据分析方法同实验 1。阅读理解选择题的平均正确率为 97.40%；语义类别选择题的平均正确率为 94.79%。表明被试在实验过程中均认真阅读了实验语句，并习得了新词的语义类别。高、低概率实验条件下对目标新词的注视情况见图 3，模型结果分析汇总见表 5。

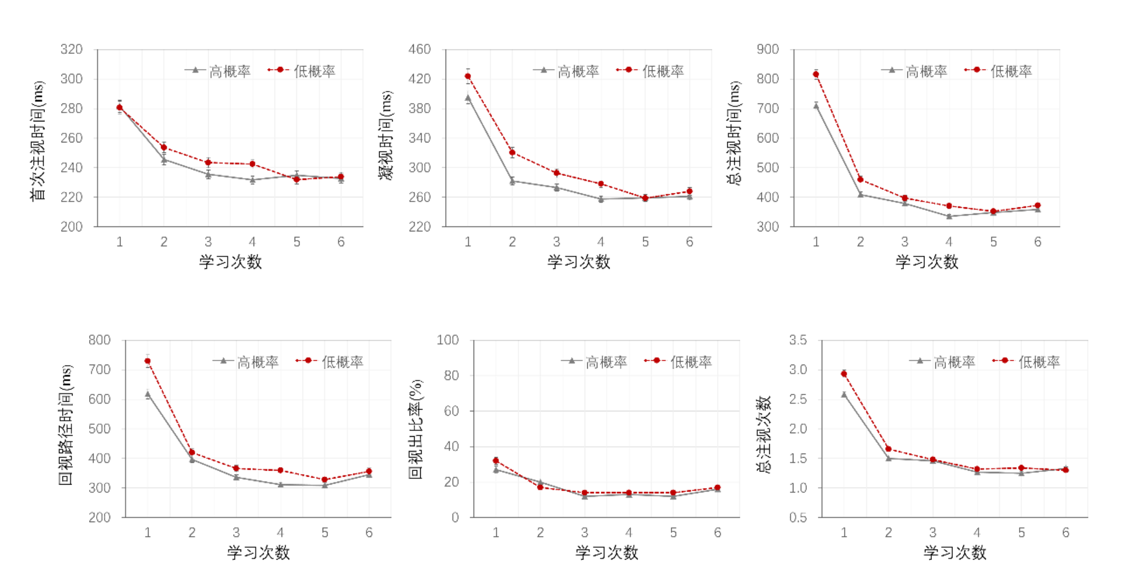


图 3 高、低尾词素位置概率条件下新词的注视情况

表 5 不同词素位置概率实验条件下的模型分析汇总结果

	变量	*b*	*SE*	*t/z*	95%CI
首次注视时间	截距	5.54	0.02	299.10	[5.50, 5.57]
	尾词素位置概率	–0.02	0.02	–0.97	[–0.05, 0.02]
	学习次数	–0.03	0.003	–8.99	[–0.03, –0.02]
	尾词素位置概率×学习次数	0.005	0.004	0.12	[–0.01, 0.01]
凝视时间	截距	5.78	0.03	229.84	[5.73, 5.83]
	尾词素位置概率	–0.08	0.02	–3.90	[–0.12, –0.04]
	学习次数	–0.06	0.005	–13.19	[–0.07, –0.05]
	尾词素位置概率×学习次数	0.01	0.01	2.14	[0.001, 0.02]
总注视时间	截距	6.30	0.05	127.52	[6.20, 6.40]
	尾词素位置概率	–0.12	0.02	–5.46	[–0.17, –0.08]
	学习次数	–0.12	0.01	–12.37	[–0.14, –0.10]
	尾词素位置概率×学习次数	0.02	0.01	3.28	[0.01, 0.03]
回视路径时间	截距	6.12	0.04	159.41	[6.05, 6.20]
	尾词素位置概率	–0.12	0.03	–4.22	[–0.18, –0.06]
	学习次数	–0.10	0.01	–12.72	[–0.12, –0.08]
	尾词素位置概率×学习次数	0.02	0.01	2.18	[0.002, 0.03]
回视出比率	截距	–1.21	0.14	–8.60	[–1.49, –0.93]
	尾词素位置概率	–0.10	0.13	–0.75	[–0.36, 0.16]
	学习次数	–0.17	0.03	–6.42	[–0.22, –0.12]
	尾词素位置概率×学习次数	–0.002	0.03	–0.07	[–0.07, 0.06]
总注视次数	截距	2.44	0.08	28.75	[2.27, 2.60]
	尾词素位置概率	–0.31	0.05	–6.46	[–0.41, –0.22]
	学习次数	–0.23	0.02	–14.08	[–0.27, –0.20]
	尾词素位置概率×学习次数	0.06	0.01	5.28	[0.04, 0.08]

在所有眼动指标分析中，学习次数的主效应均显著（$|t/z|$s ＞ 6.42, ps ＜ 0.001），随着新词学习次数的增加，被试对新词的注视时间逐步缩短，回视比率逐步降低，再次为阅读伴随词汇学习的“累积性”提供实验证据。

在首次注视时间分析中，尾词素位置概率的主效应，以及与学习次数的交互作用均不显著（$|t|$s ＜ 0.97, ps ＞ 0.05），表明在新词加工的早期阶段，读者对尾词素的位置概率信息不敏感。

在凝视时间、总注视时间、总注视次数分析中，尾词素位置概率的主效应，以及与学习次数的交互作用均显著（$|t|$s ＞ 2.14, ps ＜ 0.05）。相比于低尾词素位置概率条件，被试在高尾词素位置概率条件下对新词的凝视时间和总注视时间显著缩短，总注视次数显著减少，表现出显著的尾词素位置概率效应。进一步的交互作用分析发现（见图 4a、4b、4c），该效应随着新词学习次数的增加逐步减小，最后消失。

在回视出比率分析中，尾词素位置概率的主效应，以及与学习次数的交互作用均不显著（$|z|s < 0.75, ps > 0.05$）。但在回视路径时间分析中，尾词素位置概率的主效应，以及与学习次数的交互作用均显著（$|t|s > 2.18, ps < 0.05$，交互作用见图 4d）。被试在高尾词素位置概率条件下对新词的回视路径时间显著短于低尾词素位置概率条件，表现出显著的尾词素位置概率效应。进一步的交互作用分析发现，该效应随着新词学习次数的增加逐步减小，最后消失。上述两个与回视相关的眼动指标分析表明，随着新词在阅读中学习次数的增加，新词尾词素位置概率不会影响读者对目标词前语境的回视比率，但是会影响对目标词前语境的注视时间。

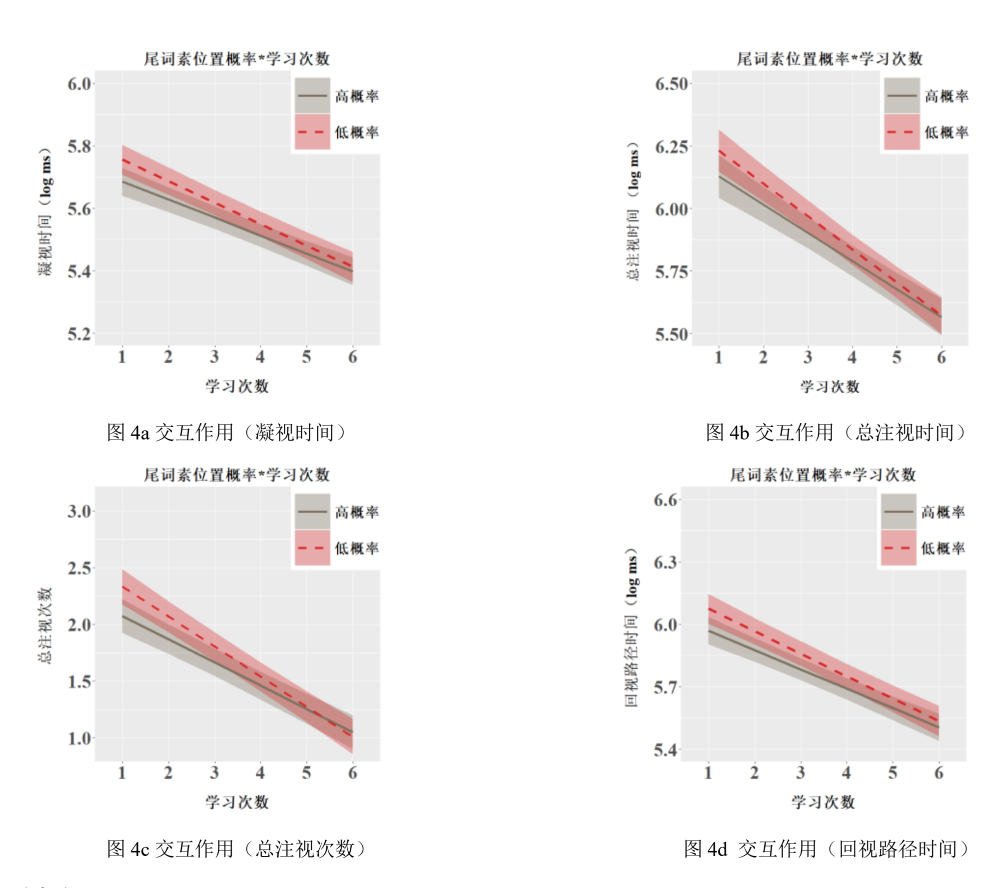


图 4a 交互作用（凝视时间）

图 4b 交互作用（总注视时间）

图 4c 交互作用（总注视次数）

图 4d 交互作用（回视路径时间）

3.3 讨论

实验 2 通过操纵新词尾词素位置概率的高低，考察了读者在阅读伴随词汇学习中是否利用尾词素的位置概率信息进行词切分。首先，与实验 1 类似，在反映词汇加工相对晚期的眼动指标中（如凝视时间、回视路径时间、总注视时间以及总注视次数），发现了显著的尾词素位置概率效应，即新词尾词素常用在词尾时，其加工时间显著短于尾词素不常用在词尾时。该发现验证了本研究的第一个假设，且与 Yen 等人(2012)和 Liang 等人(2023)的研究结论一致，表明尾词素位置概率信息作用于阅读伴随词汇学习的词切分。

本实验的第二个发现则是，在凝视时间、总注视时间、回视路径时间和总注视次数四个相对晚期眼动指标中，尾词素位置概率和学习次数的交互作用均显著。随着新词在阅读中学习次数的增加，尾词素位置概率效应逐步减小，最后消失。该发现符合本研究的第二个假设，表明尾词素位置概率信息的词切分作用同样表现出“熟悉性效应”或“学习效应”。

对比实验 1 和实验 2，除了首次注视时间这个反映词汇早期加工的眼动指标外，在词汇加工相对晚期的眼动指标（回视出比率除外）中均发现了首词素和尾词素的位置概率效应，表明首、尾词素位置概率信息均作用于中文阅读伴随词汇学习的词切分，且加工时程类似。这与泰文阅读中的研究发现一致(Kasisopa et al., 2013, 2016)。泰文作为一种无空格拼音文字语言，没有明显的视觉词切分线索。对于泰语读者而言，无论是成人还是儿童，均可利用首、尾词素位置概率信息作为一种统计学词切分线索，促进词汇识别，并引导读者将眼跳定位到词内最佳位置。

需要注意的是，词素位置概率信息的作用随学习次数的变化则表现出首、尾词素之间的差异。在实验 1 中，首词素位置概率与学习次数的交互作用表现在回视路径时间和总注视次数两个词汇加工晚期的眼动指标，并未表现在凝视时间这个反映词汇加工相对早期的眼动指标。上述结果表明在词汇加工的相对早期阶段，首词素位置概率信息作用贯穿于新词学习的全程（从第 1 次阅读至第 6 次阅读），并未表现出首词素位置概率信息的“熟悉性效应”；而在词汇加工的相对晚期阶段，首词素位置概率信息的作用随着新词学习次数的递增而逐步减小，最后消失，表现出相应的“熟悉性效应”。在实验 2 中，尾词素位置概率信息与学习次数的交互作用则从凝视时间这个反映词汇加工相对早期的眼动指标开始，一直持续到回视路径时间、总注视时间和总注视次数三个词汇加工晚期的眼动指标。表明在词汇加工的早期阶段，尾词素位置概率信息的词切分作用就开始随着新词学习次数的增多而逐步消失，表现出尾词素位置概率信息的“熟悉性效应”。从二者的交互作用图可以发现，熟悉性效应实则是在前几次新词学习中，读者利用新词首、尾词素位置概率信息进行词切分和词识别。随着新词表征的逐步构建与巩固，在后几次的新词学习中，该信息则不再起到词切分的作用。由此推断，相比于尾词素，首词素位置概率信息在阅读伴随词汇学习中的词切分作用时程更长，更稳定。该发现为首词素在双字词中的加工优势提供了新的实验证据(Ma & Li, 2015; Tsang & Zou, 2022; Wang et al., 2017; Yan et al., 2006)。

4 总讨论

本研究通过两个平行实验分别操纵首、尾词素位置概率高低，考察了首、尾词素位置概

率信息如何作用于阅读伴随词汇学习中的词切分。本研究有如下三个发现：（1）首、尾词素位置概率信息均作用于中文阅读伴随词汇学习的词切分。（2）首、尾词素位置概率信息的词切分作用表现出“熟悉性效应”，随着新词在阅读中学习次数的增加，其词切分作用逐步变小，最后消失。（3）相比于尾词素，首词素的位置概率信息在阅读伴随词汇学习中的词切分作用时程更长，更稳定。结合中文阅读在文本呈现方式上的特殊性，以及当前主流的中文阅读眼动控制模型，对上述发现进行讨论。

4.1 首、尾词素位置概率信息词切分作用的不同

本研究发现，首、尾词素位置概率信息均作用于阅读伴随词汇学习的词切分，但是该信息的词切分作用随学习次数的变化表现出首、尾词素之间的差异：在新词的前几次学习中，首、尾词素位置概率信息均起作用，在凝视时间、回视路径时间、总注视时间和总注视次数等眼动指标中均表现出首、尾词素位置概率效应；在新词的后几次学习中，首词素的位置概率信息继续起到词切分的作用（在凝视时间和总注视时间上表现出首词素位置概率效应），但与前几次阅读相比，该信息的词切分作用有减小的趋势，表现为在回视路径时间、总注视次数两个眼动指标上首词素位置概率信息和学习次数的交互作用。相比之下，在新词的后几次学习中，尾词素位置概率信息的词切分作用则完全消失。上述研究结果表明，首、尾词素位置概率信息在阅读伴随词汇学习中的作用方式受新词学习阶段的调节。

依据复合词的混合通达表征模型(Caramazza et al., 1988)，在新词学习的早期阶段，新词的加工方式类似于低频词（在心理词典中倾向于以词素形式储存），其识别受词素表征影响较大，使得首、尾词素固有的位置概率信息得到激活，帮助词切分和词识别；在新词学习的后期阶段，新词的加工方式开始逐步向中频词，甚至高频词转变（在心理词典中倾向于以整词形式储存），整词表征在词汇识别中所起的作用越来越大，相应地，词素表征在词汇识别中的作用则越来越小，此时，读者对词素水平的位置概率信息敏感性降低，使得首、尾词素位置概率信息的词切分作用逐步减小。

综合以往文献，在阅读中完全习得一个新词，至少需要在阅读中出现 12~15 次(Joseph et al., 2014; Liang et al., 2021; Nation et al., 2007; Tamura et al., 2017)。本研究中的新词在阅读中仅出现 6 次，距离形成完整的新词表征尚远，未能达到高频词的表征方式，充其量只能算是一个低－中频词。此时，首词素的词切分作用依然存在，而尾词素的词切分作用则消失。结合本研究结果，在新词学习的后几次，首词素位置概率信息的词切分作用依然存在，而尾词素位置概率信息的词切分作用消失，表明随着新词学习的深入，尾词素的词切分作用首先消

失，然后再是首词素词切分作用的消失。由此推断，在中文阅读伴随词汇学习过程中，相比于尾词素，首词素位置概率信息所起的词切分作用时程更长。当读者继续在阅读中学习同一新词时，随着新词表征的巩固与完善，新词的加工方式越来越接近高频词。当整词表征占主导时，将重复曹海波等人(2023)的研究成果，首、尾词素位置概率信息的词切分作用均消失。

由此推断，在中文词汇习得的全程中，词素位置概率信息的词切分作用随着学习的深入发生变化，见图 5 所示：在新词学习的早期，首、尾词素位置概率信息均起作用；随着新词学习的深入，尾词素位置概率信息的作用首先递减，再是首词素位置概率信息作用的递减；在新词学习的后期，首、尾词素位置概率信息的词切分作用均消失。既然在新词学习的早期阶段，首、尾词素位置概率信息均起到词切分的作用，那么，后续研究有必要明确二者的共同作用方式。此外，后续研究也有必要明确首、尾词素位置概率的词切分作用消失的时间与新词表征构建程度之间的关系，即当新词学习到什么程度时读者就不再依据首、尾词素的位置概率信息进行词切分与词识别。

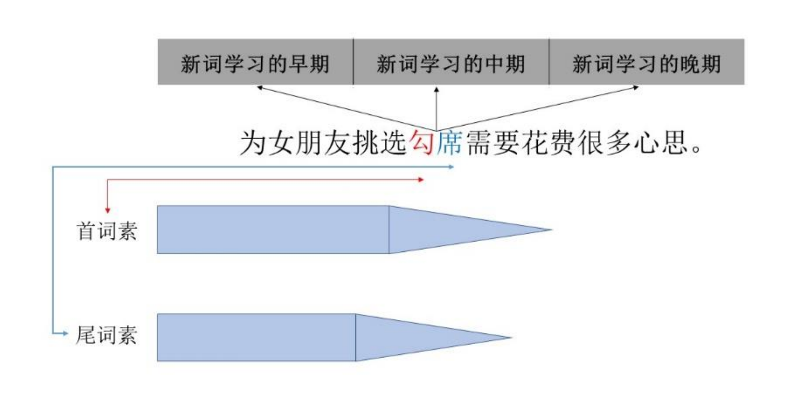


图 5 词素位置概率信息词切分作用随学习的深入发生变化的模式

注：蓝色阴影大小表示首、尾词素位置概率词切分作用大小

首、尾词素位置概率信息在词切分作用中的差异，一方面为中文双字词（特别是低频词）识别中首词素具有加工优势提供了新的实验证据(曹海波 等, 2023; Ma & Li, 2015; Milledge et al., 2022; Tsang & Zou, 2022; Wang et al., 2017; Yan et al., 2006)；一方面也证实了基于拼音文字阅读提出的自我组织词汇习得与识别模型以及顺序编码模型所提出的核心假设的正确性——词汇中字母激活程度由词首向词尾逐步递减(Davis, 2001; Whitney, 2001)。虽然中文的词长较拼音文字语言（如英语、芬兰语等）变异程度较小，以双字词为主，系列实验证据依然在中文阅读中肯定了首词素的加工优势。这可能是由于：（1）中文文本从左到右的书写和

阅读加工方式使得首词素更为重要；（2）从左到右的视觉处理方式使得首词素的信息更易获得；（3）只有从左至右依次加工首、尾词素，才能获得词汇的语音信息。基于上述实验证据，有必要在中文阅读眼动控制模型的发展与完善中，将首词素的加工优势纳入模型，用以增强模型的解释力。

4.2 首、尾词素位置概率信息加工对于理解中文阅读词切分机制的启示

由于中文阅读无词间空格之类的视觉信息作为词切分线索，其词切分机制相对复杂。新近的中文阅读眼动控制模型主张，中文的词切分和词识别是同一过程，当前词被切分出来的同时就意味着词汇已经被识别(Li & Pollatsek, 2020)。Liang 等人(2023)基于该观点以及她们研究团队的发现，解释了为何尾词素，而不是首词素的位置概率信息在中文阅读中起到词切分作用。当读者开始阅读一句话时（如“快乐阅读是我们最美的教育追求”），句首第一个词的左边界是确定的，读者只要依据尾词素（“乐”）的位置概率信息判断首词的结尾在哪里，就能完成首词的切分与识别。由于中文阅读中相邻两个词共享一个词边界，上一词的词尾（“乐”）与下一词的词首（“阅”）共享一个边界，即上一词（“快乐”）的右边界就是下一词（“阅读”）的左边界，因此，当读者完成前一个词的切分与识别时，就意味着下一个词的词首位置同时被识别。那么，基于认知加工的经济性原则，读者无需再利用首词素的位置概率信息进行词首的再切分，而只需利用尾词素的位置概率信息判断词尾在哪里即可完成当前词识别。同时，下一词的词首位置已经被识别，依此类推，完成句子阅读与理解。

显然，Liang 等人(2023)目前的观点无法解释本研究的发现。本研究所发现的首、尾词素位置概率信息的加工方式受词频所调节，对于发展与修正该理论解释有如下启示：首词素位置概率信息是否起作用，在一定程度上依赖于词汇的加工难度，这可能与当前词切分正确与否的检验过程有关。进一步解释为，虽然词首位置在识别上一词时就已确定，但是由于高频词的加工相对容易，读者可以在副中央凹处利用下一词更多的预视信息，进行首次词切分，使得首次切分正确率相对较高。此时，读者就不再需要激活首词素的位置概率信息完成词切分的检查过程；相比之下，由于低频词或新词的加工相对较难，虽然词首位置在识别上一词时也已确定，但由于读者在进行首次切分时所能利用的下一词的预视信息相对较少，导致首次切分的正确率相对较低。在完成当前词切分正确与否的检验过程时，读者可能会再次激活首词素的位置概率信息。基于上述讨论，有必要在后续研究中进一步考察首词素位置概率信息作用于中文阅读的发生条件，在此基础上理解中文阅读文本呈现方式与首词素位置概率信息词切分作用的权衡。

首、尾词素位置概率信息的加工方式受词频所调节这一研究结论，为 Yu 等人(2021)所主张的基于字词熟悉性的词切分算法提供了直接的实验证据。如前所述，词素位置概率从本质上是基于汉字位于词内特定位置的构词力的统计信息，可能在一定程度上影响词汇的熟悉性计算。如果一个汉字位于词首所构成的双字词数量较多，且使用频率较高，则意味着该汉字的熟悉性相对较高，切分起来相对容易；反之，会增加词切分难度。后续研究有必要明确词素位置概率信息与字、词熟悉性之间的关系，并尝试将其纳入模型，解释中文阅读中基于字词熟悉性计算的词切分机制。

5 结论

本研究条件下得出如下结论：（1）首、尾词素位置概率信息均作用于阅读伴随词汇学习的词切分。（2）相比于尾词素，首词素的位置概率信息在阅读伴随词汇学习中的词切分作用时程更长，更稳定。

参考文献

Bai, X. J., Liang, F. F., Blythe, H. I., Zang, C. L., Yan, G. L., & Liversedge, S. P. (2013). Interword spacing effects on the acquisition of new vocabulary for readers of Chinese as a second language. *Journal of Research in Reading, 36*(S1), S4–S17.

Bai, X. J., Ma, J., Li, X., Lian, K. Y., Tan, K., Yang, Y., & Liang, F. F. (2019). The efficiency and improvement of novel word’s learning in Chinese children with developmental dyslexia during natural reading. *Acta Psychologica Sinica, 51*(4), 471–483.

[白学军, 马杰, 李馨, 连坤予, 谭珂, 杨宇, 梁菲菲. (2019). 发展性阅读障碍儿童的新词习得及其改善. *心理学报, 51*(4), 471–483.]

Bai, X. J., Yan, G. L., Liversedge, S. P., Zang, C. L., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1277−1287.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2023). *lme4: Linear mixed–effects models using 'Eigen' and S4*. Retrieved July 4, 2023, from https://cran.r-project.org/web/packages/lme4/index.html

Blythe, H. I., Liang, F. F., Zang, C. L., Wang, J. X., Yan, G. L., Bai, X. J., & Liversedge, S. P. (2012). Inserting spaces into Chinese text helps readers to learn new words: An eye movement study. *Journal of Memory & Language, 67*(2), 241–254.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX–CH: Chinese word and character frequencies based on film subtitles. *PloS One, 5*(6), e10729.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*(3), 297–332.

Cao, H. B., Lan, Z. B., Gao, F., Yu, H. T., Li, P., & Wang, J. X. (2023). The role of character positional frequency on word recognition during Chinese reading: Lexical decision and eye movements studies. *Acta Psychologica Sinica, 55*(2), 159–176.

[曹海波, 兰泽波, 高峰, 于海涛, 李鹏, 王敬欣. (2023). 词素位置概率在中文阅读中的作用:词汇判断和眼动研究. *心理学报, 55*(2), 159–176.]

Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016).

Eye movements in reading and information processing: Keith Rayner’s 40 year legacy. *Journal of Memory and Language, 86*, 1–19.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Davis, C. J. (2001). The self–organising lexical acquisition and recognition (SOLAR) model of visual word recognition. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(1–B), 594.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology, 166*, 190–211.

Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition, 133*(1), 238–248.

Kasisopa, B., Reilly, R. G., Luksaneeyanawin, S., & Burnham, D. (2013). Eye movements while reading an unspaced writing system: The case of Thai. *Vision Research, 86*, 71–80.

Kasisopa, B., Reilly, R. G., Luksaneeyanawin, S., & Burnham, D. (2016). Child readers’ eye movements in reading Thai. *Vision Research, 123*, 8–19.

Li, X. S., Huang, L. J. Q., Yao, P. P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, *1*, 133–144.

Li, X. S., & Pollatsek, A. (2020). An integrated model of word processing and eye–movement control during Chinese reading. *Psychological Review, 127*(6), 1139–1162.

Li, X. S., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58*(4), 525–552.

Lian, K. Y., Ma, J., Wei, L., Zhang, S. W., & Bai, X. J. (2013). The role of character positional frequency on college and primary student in oral reading. *Studies of Psychology and Behavior, 19*(2), 179–185.

[连坤予, 马杰, 魏玲, 张书纬, 白学军. (2021). 汉语朗读中词素位置概率线索作用的发展研究. *心理与行为研究, 19*(2), 179–185.]

Liang, F. F., Blythe, H. I., Bai, X. J., Yan, G. L., Li, X., Zang, C. L., & Liversedge, S. P. (2017). The role of character positional frequency on Chinese word learning during natural reading. *PloS One, 12*(11), e0187656.

Liang, F. F., Blythe, H. I., Zang, C. L., Bai, X. J., Yan, G. L., & Liversedge, S. P. (2015). Positional character frequency and word spacing facilitate the acquisition of novel words during Chinese children's reading. *Journal of Cognitive Psychology, 27*(5), 594–608.

Liang, F. F., Gao, Q., Li, X., Wang, Y. S., Bai, X. J., & Liversedge, S. P. (2023). The importance of the positional probability of word final (but not word initial) characters for word segmentation and identification in children and adults' natural Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(1), 98–115.

Liang, F. F., Ma, J., Bai, X. J., & Liversedge, S. P. (2021). Initial landing position effects on Chinese word learning in children and adults. *Journal of Memory and Language, 116*(1), 104183.

Liang, F. F., Ma, J., Li, X., Lian, K. Y., Tan, K., & Bai, X. J. (2019). Saccadic targeting deficits of Chinese children with developmental dyslexia: Evidence from novel word learning in reading. *Acta Psychologica Sinica, 51*(7), 805–815.

[梁菲菲, 马杰, 李馨, 连坤予, 谭珂, 白学军. (2019). 发展性阅读障碍儿童阅读中的眼跳定位缺陷:基于新词学习的实验证据. *心理学报, 51*(7), 805–815.]

Ma, G. J., & Li, X. S. (2015). How character complexity modulates eye movement control in Chinese reading. *Reading and Writing, 28* (6), 747–761.

Milledge, S. V., Liversedge, S. P., & Blythe, H. I. (2022). The importance of the first letter in children’s parafoveal

preprocessing in English: Is it phonologically or orthographically driven? *Journal of Experimental Psychology: Human Perception and Performance, 48*(5), 427–442.

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self–teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*(1), 71–84.

Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science, 43*(1), e12705.

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research, 49*(15), 1994–2000.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology, 16*(1–2), 3–26.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K. (2009). The thirty fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

R Development Core Team. (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition, 163*, 93–102.

Tsang, Y.-K., & Zou, Y. (2022). An ERP megastudy of Chinese word recognition. *Psychophysiology, 59*(11), e14111.

Wang, W. N., Lu, A. T., He, D. P., Zhang, B., & Zhang, J. X. (2017). ERP evidence for Chinese compound word recognition: Does morpheme work all the time? *Neuroquantology, 5*(3), 142–152.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: the seriol model and selective literature review. *Psychonomic Bulletin & Review, 8*(2), 221–243.

Yan, G. L., Tian, H. J., Bai, X. J., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology, 97*(2), 259–268.

Yen, M. -H., Radach, R., Tzeng, J. L., & Tsai, J. L. (2012). Usage of statistical cues for word boundary in reading Chinese sentences. *Reading and Writing, 25*(5), 1007–1029.

Yu, L. L., Liu, Y. P., & Reichle, E. D. (2021). A corpus-based versus experimental examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General, 150*(8), 1612–1641.

Zang, C. L., Liang, F. F., Bai, X. J., Yan, G. L., & Liversedge, S. P. (2013). Inter–word spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology Human Perception & Performance, 39*(3), 720–734.

Different Roles of Initial and Final Character Positional Probabilities on Incidental Word Learning during Chinese Reading

LIANG Feifei[1, 2, 3], FENG Linlin[2], LIU Ying[2], LI Xin[1, 2, 3], BAI Xuejun[1, 2, 3]

([1]*Key Research Base of Humanities and Social Sciences of the Ministry of Education, Academy of Psychology and Behavior, Tianjin Normal University, Tianjin 300387*)

([2]*Faculty of Psychology, Tianjin Normal University, Tianjin 300387*)

([3]*Tianjin Social Science Laboratory of Students' Mental Development and Learning, Tianjin 300387*)

Abstract

In natural unspaced Chinese reading, there are no salient visual word segmentation cues (like word spaces) to demark where words begin or end, yet Chinese skilled readers process a comparable amount of text content as efficiently as English readers, processing roughly 400 characters (equal to 260 words) per minute (see Liversedge et al., 2016). This raises the question of how Chinese readers engage in such word segmentation processing efficiently and effectively. Liang et al (2015, 2017) have shown that the positional probability information associated with a character, might offer a cue to the likely positions of word boundaries during Chinese incidental word learning. Given that they simultaneously manipulated the positional probabilities of both word initial and word final characters to make their manipulations maximally effective, it is unclear whether the initial, the final, or both constituent characters' positional probabilities contribute to the word segmentation and word identification effects during incidental word learning in Chinese reading. For this reason, in the present study, two parallel experiments were designed to directly investigate whether word initial, or word ending characters are more or less important for word segmentation word learning in Chinese reading.

Two-character pseudowords were constructed as novel words. Each novel word was embedded into six high-constraint contexts for readers to establish novel lexical representation. In Experiment 1, we examined how word's initial character positional probability influenced word segmentation and word identification during Chinese word learning. The initial character's positional probability of target words was manipulated as being either high or low, and the final character was kept identical across the two conditions. In Experiment 2, an analogous manipulation was made for the final character of the target word to check whether the final character positional probability of two-character words can be used as word segmentation cue. We also included "Exposure" as a continuous variable into the model to further examine how the process of initial and final character positional probabilities changed with exposure.

In both experiments, the participants spent shorter reading times and made fewer fixations on targets that comprised initial and final characters with high relative to low positional probabilities, suggesting that the positional probability of both the initial and final character of a word influences segmentation commitments in novel word learning in Chinese reading. Furthermore, both the effect of initial and final character positional probabilities of novel words decreased with exposure, showing the typical familiarity effect. To be somewhat different, the familiarity effect associated with the initial character had a slower time course relative to final character. This finding suggests that the role of word's initial character positional probability is of more importance than that of final

character’s, supporting the concurrent standpoint that word beginning constituents might be more influential than word final constituents during two-character word identification in Chinese reading.

Based on the findings above, the time course of the process of initial and final character positional probabilities of novel words is argued and summarized as follows. During the early stage of word learning, both the statistical properties of word’s initial and final character positional probabilities are processed as segmentation cue. As lexical familiarity increases, the extent to such segmentation roles decreases, which initially begins with final character, and then occurs with initial character. Later, both the roles of initial and final character positional probabilities disappear with the establishment of a more-integral representation of novel words.

Keywords character positional probability, word segmentation, incidental word learning, Chinese reading